Bayesian updating of a probability distribution encoded on a quantum register

Andrei N. Soklakov and Rüdiger Schack

Department of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom

15 November 2005

Abstract

We investigate the problem of Bayesian updating of a probability distribution encoded in the quantum state of n qubits. The updating procedure takes the form of a quantum algorithm that prepares the quantum register in the state representing the posterior distribution. Depending on how the prior distribution is given, we describe two implementations, one probabilistic and one deterministic, of such an algorithm in the standard model of a quantum computer.

1 Introduction

Bayes's rule provides a simple and fundamental mechanism for updating a probability distribution in the light of new data [1]. The rule takes its simplest form for a finite sample space, $\mathbb{H}$, where the elements $h \in \mathbb{H}$ can be identified with the atomic events, or *hypotheses*. Let $P_{\text{prior}}(h) = P(h)$ be the prior probability distribution, and assume some piece of data, d, is observed. If $P(d|h)$ is the conditional probability of d, given h, Bayesian updating consists of replacing the prior with the posterior distribution, $P_{\text{posterior}} = P(h|d)$, where

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_h P(d|h)P(h)} \,. \tag{1}$$

To simplify the notation, we assume from now on that the set of hypotheses is of the form $\mathbb{H} = \{0, \ldots, 2^n - 1\}$ for some positive integer n. For $h \in \mathbb{H}$, let $|h\rangle$ denote the computational basis states of a register of n qubits. The state

$$|\Psi_{\text{prior}}\rangle = \sum_{h \in \mathbb{H}} \sqrt{P(h)}\, |h\rangle \tag{2}$$

provides an encoding of the prior on the quantum register. Even though the size of the sample space grows exponentially with the number of qubits, n, there exists an interesting class of priors for which $|\Psi_{\text{prior}}\rangle$ can be prepared efficiently, in the sense that the required computational resources grow only polynomially with n [2, 3].

To formulate the problem of Bayesian updating for a prior encoded on a quantum register, we make the assumption that we have a classical algorithm that computes, as a function of h, the conditional probability $P(d|h)$ for the observed data d. Given this classical algorithm, the goal of Bayesian updating is then to prepare the register in the state

$$|\Psi_{\text{posterior}}\rangle = \sum_{h \in \mathbb{H}} \sqrt{P(h|d)}\, |h\rangle \ , \tag{3}$$

with $P(h|d)$ given by Eq. (1). If the prior is given to us in the form of a single copy of the state $|\Psi_{\text{prior}}\rangle$, our problem is equivalent to finding a quantum operation, M_d, that maps any prior state of the form (2) into the corresponding posterior state of the form (3),

$$M_d |\Psi_{\text{prior}}\rangle = |\Psi_{\text{posterior}}\rangle \ . \tag{4}$$

It is easy to see that M_d cannot in general be a trace-preserving map. For example, consider the two prior states

$$|\Psi^1_{\text{prior}}\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)\,, \quad |\Psi^2_{\text{prior}}\rangle = \frac{1}{\sqrt{2}}(|2\rangle + |3\rangle)\,, \tag{5}$$

corresponding to two different prior probability distributions, and assume that the conditional probability distribution is given by

$$P(d|h) = \begin{cases} 0 & \text{if } h = 2\,, \\ c \neq 0 & \text{otherwise}\,, \end{cases} \tag{6}$$

where c is a constant determined by normalization. Although the prior states (5) are nonorthogonal, we obtain mutually orthogonal posterior states

$$|\Psi^1_{\text{posterior}}\rangle = M_d |\Psi^1_{\text{prior}}\rangle = |1\rangle \ , \quad |\Psi^2_{\text{posterior}}\rangle = M_d |\Psi^2_{\text{prior}}\rangle = |3\rangle \ , \tag{7}$$

which implies that M_d is trace-decreasing. Bayesian updating of a single copy of $|\Psi_{\rm prior}\rangle$ is therefore generally probabilistic. Section II of this paper discusses probabilistic Bayesian updating.

A deterministic updating scheme is possible, however, if the prior is given in the form of a unitary quantum circuit that maps a standard state, assumed for simplicity to be the computational basis state $|0\rangle$, to $|\Psi_{\rm prior}\rangle$. Deterministic updating is the topic of Section III.

2 Probabilistic algorithms

As we have shown above there is in general no trace preserving quantum operation that can transform all prior states into the corresponding posterior state. To realize probabilistic Bayesian updating, we proceed as follows. Define

$$E_1 = C \sum_{h\in\mathbb{S}_{\rm pr}} \sqrt{P(d|h)}\,|h\rangle\langle h|\,, \tag{8}$$

where C is a constant and $\mathbb{S}_{\rm pr}$ is a set containing the support of the prior probability distribution. We see that

$$E_1|\Psi_{\rm prior}\rangle \propto |\Psi_{\rm posterior}\rangle\,. \tag{9}$$

For sufficiently small $|C|$, see Eq. (15) below, one can view E_1 as an element of a trace preserving quantum operation $\mathcal{E}$ defined, for arbitrary ρ, by

$$\mathcal{E}(\rho) = \sum_{k=0}^{1} E_k \rho E_k^\dagger = \sum_{k=0}^{1} p_k \rho(k)\,, \tag{10}$$

where

$$p_k = {\rm Tr}(E_k\rho E_k^\dagger) \quad \text{and} \quad \rho(k) = E_k\rho E_k^\dagger/p_k\,. \tag{11}$$

This decomposition shows that the operation $\mathcal{E}$ can be realized as a measurement with outcomes $k = 0, 1$, where each outcome k happens with probability p_k and the corresponding conditional density matrix is $\rho(k)$. Substituting $\rho = |\Psi_{\rm prior}\rangle\langle\Psi_{\rm prior}|$ we see that the measurement outcome 1 corresponds to successful Bayesian updating. This happens with probability

$$p_1 = \langle\Psi_{\rm prior}|E_1^\dagger E_1|\Psi_{\rm prior}\rangle = C^2 \sum_h P(h)P(d|h) = C^2 P(d)\,. \tag{12}$$

In order to obtain a bound on C, we note that

$$E_0^\dagger E_0 = \mathbb{1} - E_1^\dagger E_1 = \mathbb{1} - C^2 \sum_{h \in \mathbb{S}_{\rm pr}} P(d|h)\, |h\rangle\langle h| \,. \tag{13}$$

Using the positivity of $E_0^\dagger E_0$, we find

$$C^2 \leq \left(\sum_{h \in \mathbb{S}_{\rm pr}} P(d|h)\, |\langle v|h\rangle|^2 \right)^{-1} \tag{14}$$

for any vector $|v\rangle$.

Now let h^* be such that $P(d|h^*) = \max_{h \in \mathbb{S}_{\rm pr}} P(d|h)$. Since the above condition is valid for any $|v\rangle$, one can choose $|v\rangle = |h^*\rangle$ and obtain

$$C^2 \leq 1/\max_{h \in \mathbb{S}_{\rm pr}} P(d|h) \,. \tag{15}$$

Together with Eq. (12) this gives an upper bound on the success probability of Bayesian updating

$$p_1 \leq \frac{P(d)}{\max_{h \in \mathbb{S}_{\rm pr}} P(d|h)} \,. \tag{16}$$

In the next subsection we describe an explicit algorithm that achieves this bound.

2.1 Explicit algorithm

The operation $\mathcal{E}$ can be realized as a modification of a procedure proposed by Rudolph [4] as follows. First we prepare the product of the prior state and an auxiliary qubit state, $|\Psi_{\rm prior}\rangle|0\rangle$. Then, using the classical algorithm for computing $P(d|h)$, one can construct a quantum circuit U_d that performs a conditional rotation of an auxiliary qubit so that

$$U_d|\Psi_{\rm prior}\rangle|0\rangle = \sum_h \sqrt{P(h)}|h\rangle\Big(A_1(h)|0\rangle + B_1(h)|1\rangle\Big) \,, \tag{17}$$

where

$$A_1(h) = c_1\sqrt{P(d|h)}, \quad B_1^2 = 1 - A_1^2 = 1 - c_1^2 P(d|h) \,, \tag{18}$$

and c_1 is a constant. Then measuring the auxiliary qubit we obtain the desired state $|\Psi_{\rm posterior}\rangle|0\rangle$ with probability

$$p_1 = c_1^2 \sum_h P(h)P(d|h) = c_1^2 P(d)\,. \tag{19}$$

Looking at Eqs.(17) and (18) we can set $c_1^2 = 1/\max_{h\in\mathbb{S}_{\rm pr}} P(d|h)$. With this setting, p_1 achieves the theoretical bound on the success probability, Eq.(16).

In the above algorithm, one can safely achieve the maximal success probability only if the knowledge of the value of $\max_{h\in\mathbb{S}_{\rm pr}} P(d|h)$ is available. It is relevant to mention here that the lack of such knowledge does not prevent us from using the above algorithm, since we can always use the trivial setting $c_1^2 = 1$. The price to pay is a smaller success probability.

An intermediate situation occurs if a nontrivial upper bound on $P(d|h)$ is known, i.e., a constant M such that $\max_{h\in\mathbb{S}_{\rm pr}} P(d|h) < M < 1$. One can then set $c_1^2 = 1/M$, which improves the success probability compared to the trivial setting.

2.2 Iterative algorithm

Let M_1 be an upper bound on $\max_{h\in\mathbb{S}_{\rm pr}} P(d|h)$. Imagine that at the beginning we do not have enough information about $P(d|h)$ and $P(h)$ to calculate a nontrivial value for M_1. In other words, we have to assume that $M_1 = 1$. Imagine also that we expect to acquire a better bound $M_2 < M_1$ in the future. We will now address the following question: Can we run the probabilistic algorithm of Sec. 2.1 first with the trivial bound $M_1 = 1$, and later with the improved bound M_2, without reducing the overall success probability that can be achieved by running the algorithm once with the bound M_2? We will find that this is indeed the case. This result remains true for a sequence of bounds, $M_k < M_{k-1} < \cdots < M_1$. Below we describe an iterative version of the above algorithm that makes use of better bounds as they become available.

Consider the measurement part of the algorithm of Sec. 2.1. If the measurement fails, which happens with probability $1 - p_1$, we end up with the state

$$|\psi_1\rangle = \Big(N_1 \sum_h \sqrt{P(h)}B_1(h)|h\rangle\Big)|1\rangle\,, \qquad N_1^{-2} = 1 - c_1^2 P(d)\,, \tag{20}$$

where we might have set $c_1^2 = 1/M_1$ to maximize p_1. Since we know the exact form of $|\psi_1\rangle$ we may attempt to achieve our original goal by performing a transformation

$$|\psi_1\rangle \longrightarrow N_1 \sum_h \sqrt{P(h)} B_1(h)|h\rangle \Big(A_2(h)|0\rangle + \frac{B_2(h)}{B_1(h)}|1\rangle \Big) , \tag{21}$$

where we set

$$A_2(h) = c_2 \frac{\sqrt{P(d|h)}}{B_1(h)} , \qquad B_2^2 = (1 - A_2^2) B_1^2 = B_1^2 - c_2^2 P(d|h) , \tag{22}$$

and c_2 is a constant. First of all, it is important to note that this procedure should not be attempted when c_1^2 was set to $1/M_1$, and M_1 is still the best available bound. This is because in the worst case there will be at least one hypotheses h^* which is present in the sum Eq.(21) with $B_1(h^*) = 0$ and $A_2(h^*) > 1$. It follows that the above procedure should only be applied if a better bound $M_2 > M_1$ became available (or when $c_1^2 < 1/M_1$). In this case, measurement of the auxiliary qubit yields the desired state $|\Psi_{\rm posterior}\rangle|0\rangle$ with probability

$$p_2 = N_1^2 c_2^2 \sum_h P(h) P(d|h) = \frac{c_2^2 P(d)}{1 - c_1^2 P(d)} . \tag{23}$$

Alternatively, with probability $1 - p_2$, we may end up with the state

$$|\psi_2\rangle = \Big(N_2 \sum_h \sqrt{P(h)} B_2(h)|h\rangle \Big) |1\rangle . \tag{24}$$

This state is similar in structure to the state $|\psi_1\rangle$ so we may try to recover in the same way by performing the transformation

$$|\psi_2\rangle \longrightarrow N_2 \sum_h \sqrt{P(h)} B_2(h)|h\rangle \Big(A_3(h)|0\rangle + \frac{B_3(h)}{B_2(h)}|1\rangle \Big) , \tag{25}$$

followed by the measurements of the auxiliary qubit in complete analogy with our earlier analysis. By continuing this procedure we obtain the sequence of success probabilities $p_1, p_2, \ldots$ together with the coefficients $\{A_k^2\}$ and $\{B_k^2\}$. We have

$$A_k(h) = c_k \frac{\sqrt{P(d|h)}}{B_{k-1}(h)} , \qquad B_k^2 = B_{k-1}^2 - c_k^2 P(d|h) , \tag{26}$$

and

$$p_k = \frac{c_k^2 P(d)}{\langle B_{k-2}^2 \rangle - c_{k-1}^2 P(d)}\,, \tag{27}$$

where $B_{-1}^2 = B_0^2 = 1$, $c_0^2 = 0$ and

$$\langle B_k^2 \rangle = \sum_h P(h) B_k^2(h)\,. \tag{28}$$

The constants $\{c_k\}$ are the only free parameters in this algorithm. As we have seen in the case $k = 1$, the constants $\{c_k\}$ cannot be chosen freely, and the optimal choice for them depends on the sequence $\{M_k\}$. From Eq.(26) we obtain

$$B_k^2 = 1 - P(d|h) \sum_{s=1}^{k} c_s^2 \geq 0\,, \tag{29}$$

and therefore

$$\sum_{s=1}^{k} c_s^2 \leq \frac{1}{P(d|h)}\,. \tag{30}$$

This condition must be satisfied for all h in the support of the prior and so we have

$$\sum_{s=1}^{k} c_s^2 \leq \frac{1}{\max_{h \in \mathbb{S}_{\rm pr}} P(d|h)}\,. \tag{31}$$

From Eqs. (28) and (29) we compute

$$\langle B_{k-2}^2 \rangle = 1 - P(d) \sum_{s=1}^{k-2} c_s^2\,. \tag{32}$$

Together with Eq. (27), this implies

$$p_k = \frac{P(d) c_k^2}{1 - P(d) \sum_{s=1}^{k-1} c_s^2}\,. \tag{33}$$

The probability that the algorithm is not successful after the nth stage is given by

$$P_{\rm fail}^n = \prod_{k=1}^{n} (1 - p_k) = 1 - P(d) \sum_{s=1}^{n} c_s^2\,, \tag{34}$$

which gives the corresponding success probability

$$P^n_{\rm succ} = 1 - P^n_{\rm fail} = P(d) \sum_{s=1}^{n} c_s^2 \leq P(d) / \max_{h \in \mathbb{S}_{\rm pr}} P(d|h)\,, \tag{35}$$

where we used the inequality (31). We see that the theoretical bound for the overall success probability of transforming one copy of the prior state $|\Psi_{\rm prior}\rangle$ into one copy of the posterior state $|\Psi_{\rm posterior}\rangle$ is achieved for as long as at some stage n of the algorithm we have

$$\sum_{s=1}^{n} c_s^2 = \frac{1}{\max_{h \in \mathbb{S}_{\rm pr}} P(d|h)}\,. \tag{36}$$

Given the sequence of upper bounds $M_1 > M_2 > \cdots > M_k$, and assuming that the information in the first $k-1$ of them was already used without success, the optimal value c_k^2 for the next iteration of the algorithm, which takes into account the bound M_k, can be calculated as

$$c_k^2 = \frac{1}{M_k} - \sum_{s=1}^{k-1} c_s^2 = \frac{1}{M_k} - \frac{1}{M_{k-1}}\,. \tag{37}$$

3 Deterministic updating

In this section we will assume that the prior is given in the form of a unitary quantum circuit, U, that maps the computational basis state $|0\rangle$, to the prior state. Apart from the constraint $U|0\rangle = |\Psi_{\rm prior}\rangle$, U is arbitrary. We first give an algorithm for the special case of hypothesis elimination and then show how to extend it to two-valued and more general models.

3.1 Hypothesis elimination

Imagine the situation where each piece of data d partitions the set of hypotheses $\mathbb{H}$ into two subsets: $\mathbb{H}_d$ containing all hypotheses that are consistent with d, and $\mathbb{H} \setminus \mathbb{H}_d$ containing all hypotheses that are rejected by the data d. This leads to a special case of Bayesian updating where $P(d|h)$ takes only two different values [5],

$$P(d|h) = \begin{cases} 1/|\mathbb{H}_d| & \text{if } h \in \mathbb{H}_d\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{38}$$

where $|\mathbb{H}_d|$ is the number of hypotheses that are consistent with the data d. The posterior state (3) takes the simple form

$$|\Psi_{\text{posterior}}\rangle = N \sum_{h\in\mathbb{H}_d} \sqrt{P(h)}|h\rangle\,, \tag{39}$$

where N is the normalization factor.

Using the given classical algorithm for computing $P(d|h)$, we define a quantum oracle, O_d, as

$$O_d|h\rangle = \begin{cases} -|h\rangle & \text{if } h \in \mathbb{H}_d\,, \\ |h\rangle & \text{otherwise}\,. \end{cases} \tag{40}$$

Furthermore, let Π be a conditional phase shift defined by

$$\Pi|h\rangle = \begin{cases} -|h\rangle & \text{if } h \neq 0\,, \\ |h\rangle & \text{if } h = 0\,. \end{cases} \tag{41}$$

These operations are combined with U to form an operation, $\mathcal{A}$, defined by [6]

$$\mathcal{A} = U^{-1}\Pi\, U O_d\,. \tag{42}$$

The circuit for $\mathcal{A}$ is the basic block of the quantum algorithm to prepare $|\Psi_{\text{posterior}}\rangle$.

It will be convenient to rewrite the prior state (2) in the form

$$|\Psi_{\text{prior}}\rangle = \sin\frac{\vartheta}{2}\,|\alpha\rangle + \cos\frac{\vartheta}{2}\,|\beta\rangle\,, \tag{43}$$

where

$$|\alpha\rangle = S_{\mathbb{H}_d}^{-1/2} \sum_{h\in\mathbb{H}_d} \sqrt{P(h)}\,|h\rangle\,, \qquad S_{\mathbb{H}_d} = \sum_{h\in\mathbb{H}_d} P(h)\,, \tag{44}$$

$$|\beta\rangle = S_{\mathbb{H}\setminus\mathbb{H}_d}^{-1/2} \sum_{h\in\mathbb{H}\setminus\mathbb{H}_d} \sqrt{P(h)}\,|h\rangle\,, \qquad S_{\mathbb{H}\setminus\mathbb{H}_d} = \sum_{h\in\mathbb{H}\setminus\mathbb{H}_d} P(h)\,, \tag{45}$$

and

$$\sin\frac{\vartheta}{2} = \sqrt{S_{\mathbb{H}_d}}\,. \tag{46}$$

The last equation shows that knowing the total prior probability of the hypotheses that are consistent with the data d is equivalent to knowing the value of ϑ.

It can now be shown that repeated application of the circuit $\mathcal{A}$ takes $|\Psi_{\rm prior}\rangle$ through the sequence of states

$$\mathcal{A}^k|\Psi_{\rm prior}\rangle = \sin\left(\frac{2k+1}{2}\vartheta\right)\,|\alpha\rangle + \cos\left(\frac{2k+1}{2}\vartheta\right)\,|\beta\rangle\,. \tag{47}$$

The number of times, T, of applications of $\mathcal{A}$ that achieve the required transformation,

$$\mathcal{A}^T|\Psi_{\rm prior}\rangle = |\alpha\rangle = |\Psi_{\rm posterior}\rangle\,, \tag{48}$$

is therefore

$$T = (\pi/\vartheta - 1)/2\,. \tag{49}$$

If T is not an integer, there are two possibilities. Either one uses the closest integer approximation to T and includes the effect of the noninteger part in the fidelity analysis (see below), or one follows $\lfloor T \rfloor$ applications of $\mathcal{A}$ with one application of a modified version of $\mathcal{A}$ where phases are shifted by less than $e^{i\pi}$ in both O_d and Π [7].

In order to compute the number of iterations, T, the value of ϑ must be known. To obtain ϑ, a version of the standard phase estimation algorithm [8] can be used as illustrated in Figure 1.

To calculate the effect of an error in the value of ϑ on the fidelity of the Bayesian transformation (48), we assume that there is an upper bound on the absolute error,

$$\Delta\vartheta \geq |\vartheta - \tilde{\vartheta}|\,, \tag{50}$$

where $\tilde{\vartheta}$ denotes the approximate value. With the definition $\tilde{T} = (\pi/\tilde{\vartheta} - 1)/2$, the fidelity is

$$F = |\langle\Psi_{\rm posterior}|\mathcal{A}^{\tilde{T}}|\Psi_{\rm prior}\rangle| = \sin\left(\frac{2\tilde{T}+1}{2}\vartheta\right). \tag{51}$$

Substituting $\vartheta = \tilde{\vartheta} \pm \Delta\vartheta$ and using the relation $(2\tilde{T}+1)\tilde{\vartheta} = \pi$ we obtain

$$F = \cos\left(\frac{2\tilde{T}+1}{2}\Delta\vartheta\right) = \cos\frac{\pi\Delta\vartheta}{2\tilde{\vartheta}} \geq 1 - \left(\frac{\pi\Delta\vartheta}{2\tilde{\vartheta}}\right)^2. \tag{52}$$

3.2 Two-valued models

A straightforward generalization of hypothesis elimination is provided by a two-valued conditional probability of the form

$$P(d|h) = \begin{cases} a_1 & \text{if } h \in \mathbb{H}_d\,, \\ a_2 & \text{otherwise}\,, \end{cases} \tag{53}$$

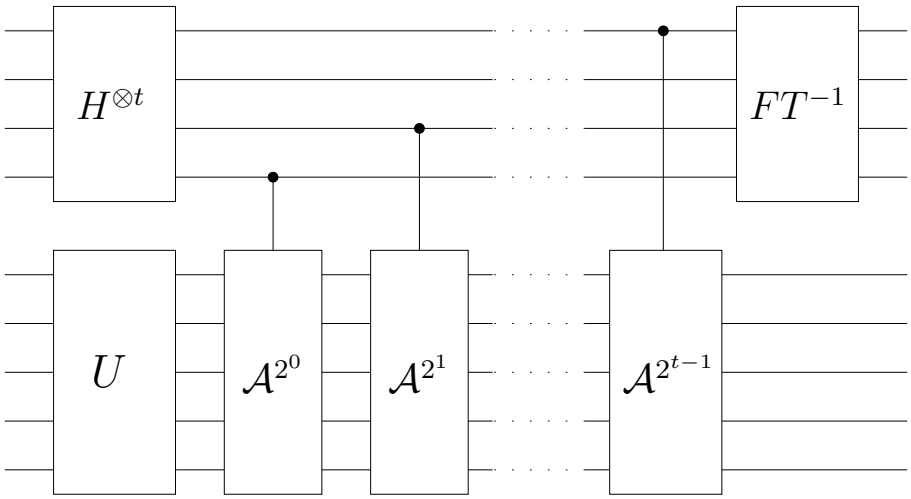

Figure 1: This is the standard phase-estimation circuit applied to the hypothesis-elimination operator $\mathcal{A}$. A measurement of the upper t-qubit register returns the value of ϑ with an accuracy of m bits and a probability of success of at least $1-\epsilon$, where m and ϵ are related to each other and to t via the condition $t = m + \lceil \log(2 + 1/2\epsilon) \rceil$. The gates labeled $H^{\otimes t}$ and FT are the t-qubit Hadamard and quantum Fourier transforms, respectively.

where $a_1 > a_2$ are constants, and $\mathbb{H}_d$ is the set of hypotheses favored by the data d. The *suppression coefficient* $r = a_1/a_2$ measures how much hypotheses in $\mathbb{H}_d$ are favored by the data. As before, the prior state can be written in the form, Eq.(43),

$$|\Psi_{\text{prior}}\rangle = \sin\frac{\vartheta}{2}\,|\alpha\rangle + \cos\frac{\vartheta}{2}\,|\beta\rangle\,, \tag{54}$$

and for the posterior state we calculate

$$|\Psi_{\text{posterior}}\rangle = \sqrt{a_1}\,\sin\frac{\vartheta}{2}\,|\alpha\rangle + \sqrt{a_2}\,\cos\frac{\vartheta}{2}\,|\beta\rangle\,. \tag{55}$$

Normalization of the posterior state implies that

$$a_2 = \frac{1}{r\sin^2(\vartheta/2) + \cos^2(\vartheta/2)}\,. \tag{56}$$

Defining ϑ' so that

$$\cos\frac{\vartheta'}{2} = \sqrt{a_2}\,\cos\frac{\vartheta}{2} = \frac{\cos(\vartheta/2)}{\sqrt{r\sin^2(\vartheta/2)+\cos^2(\vartheta/2)}}\,, \tag{57}$$

the number of iterations T necessary to transform $|\Psi_{\text{prior}}\rangle$ into $|\Psi_{\text{posterior}}\rangle = \mathcal{A}^T|\Psi_{\text{prior}}\rangle$ can then be calculated as

$$T(\vartheta, r) = (\vartheta'/\vartheta - 1)/2\,. \tag{58}$$

It follows that knowledge of ϑ and the suppression coefficient r is sufficient for a deterministic implementation of Bayesian updating with the conditional distribution (53). As before, the value of ϑ can be obtained using the algorithm of figure 1, and the same fidelity bound (52) can be used.

3.3 Bayesian updating: general models

In this section we show how to generalize the above algorithm to the case of Bayesian updating with a general model, i.e., a general conditional distribution $P(d|h)$. The main idea is to represent $P(d|h)$ as a product of two-valued models with known suppression coefficients. Bayesian updating with $P(d|h)$ can then be viewed as a sequence of Bayesian updatings for the two-valued models.

Let $C_k(h)$ be the coefficients in the binary expansion of $\log_2 P(d|h)$,

$$\log_2 P(d|h) = \sum_{k=1}^{\infty} C_k(h)\, 2^{-k}\,. \tag{59}$$

This allows us to express $P(d|h)$ as a product,

$$P(d|h) = \prod_{k=1}^{\infty} \sqrt[2^k]{2^{C_k(h)}}\,. \tag{60}$$

Let $\mathbb{H}_{d_k}$ be the set of hypotheses $\{h\}$ for which $C_k(h) = 1$. The kth term in this product is either $\sqrt[2^k]{2}$ or 1 depending on whether h is in $\mathbb{H}_{d_k}$ or not. Bayesian updating with the conditional probability $P(d|h)$ can therefore be viewed as a sequence of stages corresponding to the acquisition of data from the sequence $d_1, d_2, \ldots$. At each stage, an updating step for a two-valued model as described in the previous section is carried out.

Acknowledgments

We would like to thank Terry Rudolph for helpful discussions. This work was supported in part by the European Union IST-FET project EDIQIP.

References

[1] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, Chichester, 1994).

[2] L. Grover and T. Rudolph, e-print quant-ph/0208112.

[3] A. N. Soklakov and R. Schack, e-print quant-ph/0408045, to be published in Phys. Rev. A.

[4] T. Rudolph, private communication.

[5] A. N. Soklakov and R. Schack, e-print quant-ph/0412025.

[6] G. Brassard, P. Høyer, M. Mosca and A. Tapp, e-print quant-ph/0005055.

[7] T. Mannveille, A. N. Soklakov and R. Schack, in preparation.

[8] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).